

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: von Schaewen, Antje Dr. rer. nat.
(B) STREET: Natruperstrasse 169a
(C) CITY: Osnabrueck
(E) COUNTRY: Germany
(F) POSTAL CODE (ZIP): D-49076
(G) TELEPHONE: +49-541-684029

(ii) TITLE OF INVENTION: Plant gntI sequences and the use thereof for the production of plants having reduced or lacking N-acetyl glucosaminyl transferase I (GnTI) activity

(iii) NUMBER OF SEQUENCES: 6

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1669 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Solanum tuberosum
(B) STRAIN: Desiree
(D) DEVELOPMENTAL STAGE: Sink organ
(F) TISSUE TYPE: Mesophyll
(G) CELL TYPE: Leaf cells

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: Lambda ZAP II (Eco RI)
(B) CLONE: gntI-A1(K)

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 659..667
(D) OTHER INFORMATION:/function= "Asn codon in this context is a potential glycosylation site"
/product= "N-glycosylation consensus sequence"
/phenotype= "N-glycans modulate protein properties"
/standard_name= "N-glycosylation site"
/label= pot-CHO
/note= "GnTI-coding sequences from animals do not contain this feature"

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:53..1393
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION:/codon_start= 53
/function= "initiates complex N-glycans on
secretory glycoproteins"
/EC_number= 2.4.1.101
/product= "beta-1,2-N-acetylglucosaminyltransferase I"
/evidence= EXPERIMENTAL
/gene= "cgl"
/standard_name= "gntI"
/label= ORF
/note= "first gntI sequence from potato (unpublished)"

(ix) FEATURE:

(A) NAME/KEY: 5'UTR
(B) LOCATION:15..52

(ix) FEATURE:

(A) NAME/KEY: 3'UTR
(B) LOCATION:1394..1655

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:80..139
(D) OTHER INFORMATION:/function= "membrane anchor (amino
acids 10-29)"
/product= "hydrophobic amino acid stretch in GnTI"
/standard_name= "membrane anchor of a type II
Golgi protein"
/note= "identified by comparison with GnTI sequences
from animals"

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1..14
(D) OTHER INFORMATION:/function= "used for cloning the
cDNA library in Lambda ZAPII"
/product= "EcoRI/NotI-cDNA adapter"
/number= 1

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1656..1669
(D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAATTCCGCGG CCGCCTGAGA AACCCCTCGAA TTCAATTCG CATTGGCAG AG ATG
Met 1

AGA GGG AAC AAG TTT TGC TTT GAT TTA CGG TAC CTT CTC GTC GTG GCT
Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val Ala
5 10 15

GCT CTC GCC TTC ATC TAC ATA CAG ATG CGG CTT TTC GCG ACA CAG TCA
Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln Ser
103 151

55

103

151

20 25 30 199

GAA TAT GTA GAC CGC CTT GCT GCT GCA ATT GAA GCA GAA AAT CAT TGT
Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His Cys
35 40 45 45

ACA AGT CAG ACC AGA TTG CTT ATT GAC AAG ATT AGC CAG CAA GGA
Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gly
50 55 60 65

AGA GTA GTA GCT CTT GAA GAA CAA ATG AAG CAT CAG GAC CAG GAG TGC
Arg Val Val Ala Leu Glu Gln Met Lys His Gln Asp Gln Glu Cys
70 75 80 80

CGG CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG GGC ATA AAA
Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile Lys
85 90 95 95

AAG TTA ATC GGA GAT GTG CAG ATG CCA GTG GCA GCT GTA GTT GTT ATG
Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val Met
100 105 110 110

GCT TGC AGT CGT ACT GAC TAC CTG GAG AGG ACT ATT AAA TCC ATC TTA
Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile Leu
115 120 125 125

AAA TAC CAA ACA TCT GTT GCA TCA AAA TAT CCT CTT TTC ATA TCC CAG
Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln
130 135 140 145

GAT GGA TCA AAT CCT GAT GTA AGA AAG CTT GCT TTG AGC TAT GGT CAG
Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly Gln
150 155 160 160

CTG ACG TAT ATG CAG CAC TTG GAT TAT GAA CCT GTG CAT ACT GAA AGA
Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu Arg
165 170 175 175

CCA GGG GAA CTG GTT GCA TAC TAC AAG ATT GCA CGT CAT TAC AAG TGG
Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp
180 185 190 190

GCA TTG GAT CAG CTG TTT CAC AAG CAT AAT TTT AGC CGT GTT ATC ATA
Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile Ile
195 200 205 205

CTA GAA GAT GAT ATG GAA ATT GCT GAT TTT TTT GAC TAT TTT GAG
Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe Glu
210 215 220 225

GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG GCT ATT TCT
Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile Ser
230 235 240 240

TCT TGG AAT GAC AAT GGA CAA AGG CAG TTC GTC CAA GAT CCT GAT GCT
Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp Ala
245 250 255 255

343 391 439 487

535 583 631 679

727 775 823

CTT TAC CGC TCA GAC TTT CCT GGT CTT GGA TGG ATG CTT TCA AAA 871
Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser Lys
260 265 270
TCA ACT TGG TCC GAA CTA TCT CCA AAG TGG CCA AAG GCT TAC TGG GAT 919
Ser Thr Trp Ser Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp
275 280 285
GAC TGG CTA AGG CTG AAA GAA AAT CAC AGA GGT CGA CAA TTT ATT CGC 967
Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile Arg
290 295 305
CCA GAA GTT TGC AGA ACG TAC AAT TTT GGT GAG CAT GGT TCT AGT TTG 1015
Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu
310 315
GGG CAG TTT TTT AAG CAG TAT CTT GAG CCA ATT AAG CTA AAT GAT GTC 1063
Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val
325 330 335
CAG GTT GAT TGG AAG TCA ATG GAC CTA AGT TAC CTT TTG GAG GAC AAC 1111
Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp Asn
340 345
TAT GTG AAA CAC TTT GGC GAC TTG GTT AAA AAG GCT AAG CCC ATC CAC 1159
Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile His
355 360
GGA GCT GAT GCT GTT TTG AAA GCA TTT AAC ATA GAT GGT GAT GTG CGT 1207
Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val Arg
370 375 380 385
ATT CAG TAC AGA GAC CAA CTA GAC TTT GAA GAT ATC GCT CGA CAG TTT 1255
Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln Phe
390 395 400
GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGG GCA GCA TAT AAA 1303
Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr Lys
405 410 415
GGG ATA GTA GTT TTC CGG TTT CAA ACA TCT AGA CGT GTG TTC CTT GTT 1351
Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu Val
420 425 430
TCC CCT GAT TCT CTT CGA CAA CTT GGA GTT GAA GAT ACT TAG 1393
Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr *
435 440 445
CGAAGATATG ATTGGAGCCT GAGCAACAAT TTAGACTTAT TTGGTAGGAT ACATTTGAAA 1453
GAGCTGACAC GAAAAGTATG ACTACCAGTA GCTACATGCA ACATTTAAC GTTAATGGAA 1513
GGAACCCACT GCTTATGTT GGAATGGATG AATCATCAC ACATCCTATT ATTCAAGTTT 1573
ACAAACATAA AGAGGAAATG TTGCCCTATA AAAACAAATT TTTGTTCT AAGAAGGAAC 1633
GTTACGATTA TGAGCAACTT TGGCGGCCGC GAATTC 1669

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val
1 5 10 15
Ala Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln
20 25 30
Ser Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His
35 40 45
Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gln
50 55 60
Gly Arg Val Val Ala Leu Glu Glu Gln Met Lys His Gln Asp Gln Glu
65 70 75 80
Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile
85 90 95
Lys Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val
100 105 110
Met Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile
115 120 125
Leu Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser
130 135 140
Gln Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly
145 150 155 160
Gln Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu
165 170 175
Arg Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys
180 185 190
Trp Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile
195 200 205
Ile Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe
210 215 220
Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile
225 230 235 240
Ser Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp
245 250 255
Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser
260 265 270

Lys Ser Thr Trp Ser Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp
275 280 285
Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile
290 295 300
Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser
305 310 315 320
Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp
325 330 335
Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp
340 345 350
Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile
355 360 365
His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val
370 375 380
Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln
385 390 395 400
Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr
405 410 415
Lys Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu
420 425 430
Val Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr *
435 440 445

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1737 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Nicotiana tabacum
- (B) STRAIN: Samsun NN
- (D) DEVELOPMENTAL STAGE: Sink organ
- (F) TISSUE TYPE: Mesophyll
- (G) CELL TYPE: Leaf cells

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: Lambda ZAP II (Eco RI)
- (B) CLONE: gntI-A9(T)

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:733..741
(D) OTHER INFORMATION:/function= "Asn codon in this context is a potential glycosylation site"
/product= "N-glycosylation consensus sequence"
/phenotype= "N-glycans modulate protein properties"
/standard_name= "N-glycosylation site"
/label= pot-CHO
/note= "GnTI sequences from animals do not contain this feature"

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:127..1467
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION:/codon_start= 127
/function= "initiates complex N-glycans on secretory glycoproteins"
/EC_number= 2.4.1.101
/product= "beta-1,2-N-acetylglucosaminyltransferase I"
/evidence= EXPERIMENTAL
/gene= "cgl"
/standard_name= "gntI"
/label= ORF
/note= "first gntI sequence from tobacco (unpublished)"

(ix) FEATURE:

(A) NAME/KEY: 5'UTR
(B) LOCATION:15..126

(ix) FEATURE:

(A) NAME/KEY: 3'UTR
(B) LOCATION:1468..1723

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:154..213
(D) OTHER INFORMATION:/function= "membrane anchor (amino acids 10-29)"
/product= "hydrophobic amino acid stretch in GnTI"
/standard_name= "membrane anchor of a type II golgi protein"

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1..14
(D) OTHER INFORMATION:/function= "use for cloning the cDNA library in Lambda ZAPII"
/product= "EcoRI/NotI-cDNA adapter"
/number= 1

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1724..1737
(D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GAATTCGCGG CCGCCATTGA CTTGATCCTA ACTAACAGG CAAAGTAAAT CCAGCGATGA 60
AACACTCATA ACTAACACT GAGAGACTAT TCGCTTCCTC CTAAGCCTT CAATCGAATT 120
CGCACG ATG AGA GGG AAC AAG TTT TGC TGT GAT TTC CGG TAC CTC CTC 168
Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu
455 460
ATC TTG GCT GTC GCC TTC ATC TAC ACA CAG ATG CGG CTT TTT GCG 216
Ile Leu Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala
465 470 475
ACA CAG TCA GAA TAT GCA GAT CGC CTT GCT GCA ATT GAA GCA GAA 264
Thr Gln Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ile Glu Ala Glu
480 485 490
AAT CAT TGT ACA AGC CAG ACC AGA TTG CTT ATT GAC CAG ATT AGC CTG 312
Asn His Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Gln Ile Ser Leu
495 500 505
CAG CAA GGA AGA ATA GTT GCT CTT GAA GAA CAA ATG AAG CGT CAG GAC 360
Gln Gln Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp
510 515 520 525
CAG GAG TGC CGA CAA TTA AGG GCT CTT GAT CTT GAA AGT AAG 408
Gln Glu Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys
530 535 540
GGC ATA AAA AAG TTG ATC GGA AAT GTA CAG ATG CCA GTG GCT GCT GTA 456
Gly Ile Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val
545 550 555
GTT GTT ATG GCT TGC AAT CGG GCT GAT TAC CTG GAA AAG ACT ATT AAA 504
Val Val Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys
560 565 570
TCC ATC TTA AAA TAC CAA ATA TCT GTT GCG TCA AAA TAT CCT CTT TTC 552
Ser Ile Leu Lys Tyr Gln Ile Ser Val Ala Ser Lys Tyr Pro Leu Phe
575 580 585
ATA TCC CAG GAT GGA TCA CAT CCT GAT GTC AGG AAG CTT GCT TTG AGC 600
Ile Ser Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser
590 595 600 605
TAT GAT CAG CTG ACG TAT ATG CAG CAC TTG GAT TTT GAA CCT GTG CAT 648
Tyr Asp Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val His
610 615 620
ACT GAA AGA CCA GGG GAG CTG ATT GCA TAC TAC AAA ATT GCA CGT CAT 696
Thr Glu Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His
625 630 635
TAC AAG TGG GCA TTG GAT CAG CTG TTT TAC AAG CAT ATT TTT AGC CGT 744
Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg
640 645 650
GTT ATC ATA CTA GAA GAT GAT ATG GAA ATT GCC CCT GAT TTT TTT GAC 792
Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp
655 660 665
TTT TTT GAG GCT GGA GCT ACT CTT CTT GAC GAC GAC AAG TCG ATT ATG 840

Phe Phe Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met
670 675 680 685 888
GCT ATT TCT TCT TGG AAT GAC AAT GGA CAA ATG CAG TTT GTC CAA GAT
Ala Ile Ser Ser Trp Asn Asp Asn Gly Gln Met Gln Phe Val Gln Asp
690 695 700 936
CCT TAT GCT CTT TAC CGC TCA GAT TTT CCC GGT CTT GGA TGG ATG
Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met
705 710 715 984
CTT TCA AAA TCT ACT TGG GAC GAA TTA TCT CCA AAG TGG CCA AAG GCT
Leu Ser Lys Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala
720 725 730 1032
TAC TGG GAC GAC TGG CTA AGA CTC AAA GAG AAT CAC AGA GGT CGA CAA
Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln
735 740 745 1080
TTT ATT CGC CCA GAA GTT TGC AGA ACA TAT AAT TTT GGT GAG CAT GGT
Phe Ile Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly
750 755 760 1128
TCT AGT TTG GGG CAG TTT TTC AAG CAG TAT CTT GAG CCA ATT AAA CTA
Ser Ser Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu
770 775 780 1176
AAT GAT GTC CAG GTT GAT TGG AAG TCA ATG GAC CTT AGT TAC CTT TTG
Asn Asp Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu
785 790 795 1224
GAG GAC AAT TAC GTG AAA CAC TTT GGT GAC TTG GTT AAA AAG GCT AAG
Glu Asp Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys
800 805 810 1272
CCC ATC CAT GGA GCT GAT GTC TTG AAA GCA TTT AAC ATA GAT GGT
Pro Ile His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly
815 820 825 1320
GAT GTG CGT ATT CAG TAC AGA GAT CAA CTA GAC TTT GAA AAT ATC GCA
Asp Val Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala
830 835 840 845 1368
CGG CAA TTT GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGT GCA
Arg Gln Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala
850 855 860 1416
GCA TAT AAA GGA ATA GTA GTT TTC CGG TAC CAA ACG TCC AGA CGT GTA
Ala Tyr Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val
865 870 875 1464
TTC CTT GTT GGC CAT GAT TCG CTT CAA CAA CTC GGA ATT GAA GAT ACT
Phe Leu Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr
880 885 890 1517
TAA CAAAGATATG ATTGCAGGAG CCCGGCAAA ATTTTGACT TATTGGTAG
*
GATGCATCGA GCTGACACTA AACCATGATT TTACCAAGTTA CATAACAACGT TTTAATGTTA
TACGGAGGAG CTCACTGTT TAGTGTGAA GGGATATCGG CTTCTTAGTA TTGGATGAAT
1577 1637

CATCAACACA ACCTATTATT TTAAGTGTTC AGAACATAAA GAGGAAATGT AGCCCTGTAA 1697
AGACTATACA TGGGACCATC ATAATCGCGG CCGCGAATTC 1737

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu Ile Leu
1 5 10 15
Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala Thr Gln
20 25 30
Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ile Glu Ala Glu Asn His
35 40 45
Cys Thr Ser Gln Thr Arg Leu Leu Asp Gln Ile Ser Leu Gln Gln
50 55 60
Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp Gln Glu
65 70 75 80
Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Ile
85 90 95
Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val Val Val
100 105 110
Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys Ser Ile
115 120 125
Leu Lys Tyr Gln Ile Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser
130 135 140
Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Asp
145 150 155 160
Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val His Thr Glu
165 170 175
Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys
180 185 190
Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg Val Ile
195 200 205
Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp Phe Phe
210 215 220
Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile
225 230 235 240
Ser Ser Trp Asn Asp Asn Gly Gln Met Gln Phe Val Gln Asp Pro Tyr
245 250 255

Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser
260 265 270
Lys Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp
275 280 285
Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile
290 295 300
Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser
305 310 315 320
Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp
325 330 335
Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp
340 345 350
Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Ala Lys Pro Ile
355 360 365
His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val
370 375 380
Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala Arg Gln
385 390 395 400
Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr
405 410 415
Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val Phe Leu
420 425 430
Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr *
435 440 445

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1854 base pairs
- (B) TYPE: Nucleotide
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Arabidopsis thaliana
- (B) STRAIN: Columbia
- (D) DEVELOPMENTAL STAGE: Mature plants
- (F) TISSUE TYPE: All tissues

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: Lambda Uni-ZAP (EcoRI/XbaI) and
Lambda ACT (XbaI)
- (B) CLONE: pBSK-Ara-GntI-full #8

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1185..1193
(D) OTHER INFORMATION:/function= "Asn Codon is a potential glycosylation site"
/product= "Consensus sequence for N-glycosylation"
/phenotype= "N glycans modulate protein characteristics"
/standard_name= "N glycosylation site"
/label= pot-CHO
/note= "absent in animal GnTI sequences"

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:135..1469
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION:/codon_start= 135
/function= "initiates complex N glycans on secretory glycoproteins"
/EC_number= 2.4.1.101
/product= "beta-1,2-N-acetyl glucosaminyl transferase I"
/evidence= EXPERIMENTAL
/gene= "cgl"
/standard_name= "gntI"
/label= ORF
/note= "first gntI sequence from Arabidopsis (unpublished)"

(ix) FEATURE:

(A) NAME/KEY: 5'UTR
(B) LOCATION:19..134

(ix) FEATURE:

(A) NAME/KEY: 3'UTR
(B) LOCATION:1470..1848

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:157..215
(D) OTHER INFORMATION:/function= "membrane anchor (amino acids 8-27)"
/product= "hydrophobic amino-acid region in GnTI"
/standard_name= "membrane anchor of a Type II Golgi protein"
/note= "identified by comparison with animal GnTI sequences "

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1..18
(D) OTHER INFORMATION:/function= "for preparation of a cDNA library in Lambda ACT"
/product= "XhoI-cDNA-Adaptor"
/number= 1

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1849..1854
(D) OTHER INFORMATION:/product= "XhoI-cDNA-Adaptor"
/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGGCCA CGAAGGCCAC CGTTTTGTT ATAACGAACG ACACCGTTTC AAACAACCTC 60
CTTATTAGCT AGCTCCCTCC CGGC GGCAAA CACCAGAAGA TCCACCGCTT TTGATCTGGT 120
TGTTTGTGCGT CGAT ATG GCG AGG ATC TCG TGT GAC TTG AGA TTT CTT CTC 170
Met Ala Arg Ile Ser Cys Asp Leu Arg Phe Leu Leu 10
1 5 218
ATC CCG GCA GCT TTC ATG ATC TAC ATC CAG ATG AGG CTT TTC CAG
Ile Pro Ala Ala Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln 25
15 20 266
ACG CAA TCA CAG TAT GCA GAT CGC CTC AGT TCC GCT ATC GAA TCT GAG
Thr Gln Ser Gln Tyr Ala Asp Arg Leu Ser Ser Ala Ile Glu Ser Glu 30
35 40 314
AAC CAT TGC ACT AGT CAA ATG CGA GGC CTC ATA GAT GAA GTT AGC ATC
Asn His Cys Thr Ser Gln Met Arg Gly Leu Ile Asp Glu Val Ser Ile 60
45 50 362
AAA CAG TCG CGG ATT GTT GCC CTC GAA GAT ATG AAG AAC CGC CAG GAC
Lys Gln Ser Arg Ile Val Ala Leu Glu Asp Met Lys Asn Arg Gln Asp 75
65 70 410
GAA GAA CTT GTG CAG CTT AAG GAT CTA ATC CAG ACG TTT GAA AAA AAA
Glu Glu Leu Val Gln Leu Lys Asp Leu Ile Gln Thr Phe Glu Lys Lys 80
85 90 458
GGA ATA GCA AAA CTC ACT CAA GGT GGA CAG ATG CCT GTG GCT GTA
Gly Ile Ala Lys Leu Thr Gln Gly Gln Met Pro Val Ala Ala Val 105
95 100 506
GTG GTT ATG GCC TGC AGT CGT GCA GAC TAT CTT GAA AGG ACT GTT AAA
Val Val Met Ala Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys 120
110 115 554
TCA GTT TTA ACA TAT CAA ACT CCC GTT GCT TCA AAA TAT CCT CTA TTT
Ser Val Leu Thr Tyr Gln Thr Pro Val Ala Ser Lys Tyr Pro Leu Phe 140
125 130 602
ATA TCT CAG GAT GGA TCT GAT CAA GCT GTC AAG AGC AAG TCA TTG AGC
Ile Ser Gln Asp Gly Ser Asp Gln Ala Val Lys Ser Lys Leu Ser 155
145 150 650
TAT AAT CAA TTA ACA TAT ATG CAG CAC TTG GAT TTT GAA CCA GTG GTC
Tyr Asn Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val Val 170
160 165 698
ACT GAA AGG CCT GGT GAA CTG ACT GCG TAC TAC AAG ATT GCA CGT CAC
Thr Glu Arg Pro Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His 185
175 180 746
TAC AAG TGG GCA CTG GAC CAG TTG TTT TAC AAA CAC AAA TTT AGT CGA
Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Lys Phe Ser Arg 200
190 195 794
GTG ATT ATA CTA GAA GAC GAT ATG GAA ATT GCT CCA GAC TTC TTT GAT
Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp 220
205 210 215

TAC TTT GAG GCT GCA GCT AGT CTC ATG GAT AGG GAT AAA ACC ATT ATG 842
Tyr Phe Glu Ala Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met 235
225
GCT GCT TCA TCA TGG AAT GAT AAT GGA CAG AAG CAG TTT GTG CAT GAT 890
Ala Ala Ser Ser Trp Asn Asn Gly Gln Lys Gln Phe Val His Asp 250
240
CCC TAT GCG CTA TAC CGA TCA GAT TTT CCT GGC CTT GGG TGG ATG 938
Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Pro Gly Leu Gly Trp Met 265
255
CTC AAG AGA TCG ACT TGG GAT GAG TTA TCA CCA AAG TGG CCA AAG GCT 986
Leu Lys Arg Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala 280
270
TAC TGG GAT GAT TGG CTG AGA CTA AAG GAA AAC CAT AAA GGC CGC CAA 1034
Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln 300
285
TTC ATT GCA CCG GAA GTC TGT AGA ACA TAC AAT TTT GGT GAA CAT GGG 1082
Phe Ile Ala Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly 315
305
TCT AGT TTG GGA CAG TTT TTC AGT CAG TAT CTG GAA CCT ATA AAG CTA 1130
Ser Ser Leu Gly Gln Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu 330
320
AAC GAT GTG ACG GTT GAC TGG AAA GCA AAG GAC CTG GGA TAC CTG ACA 1178
Asn Asp Val Thr Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr 345
335
GAG GGA AAC TAT ACC AAG TAC TTT TCT GGC TTA GTG AGA CAA GCA CGA 1226
Glu Gly Asn Tyr Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg 360
355
CCA ATT CAA GGT TCT GAC CTT GTC TTA AAG GCT CAA AAC ATA AAG GAT 1274
Pro Ile Gln Gly Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp 380
365
GAT GAT CGT ATC CGG TAT AAA GAC CAA GTA GAG TTT GAA CGC ATT GCA 1322
Asp Asp Arg Ile Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala 395
385
GGG GAA TTT GGT ATA TTT GAA GAA TGG AAG GAT GGT GTG CCA CGA ACA 1370
Gly Glu Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr 410
400
GCA TAT AAA GGA GTA GTG GTG TTT CGA ATC CAG ACA ACA AGA CGT GTA 1418
Ala Tyr Lys Gly Val Val Phe Arg Ile Gln Thr Arg Arg Val 425
415
TTC CTG GTT GGG CCA GAT TCT GTA ATG CAG CTT GGA ATT CGA AAT TCC 1466
Phe Leu Val Gly Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser 440
430
TGA TGC~~AAAACAT~~ ATGAAAGGAA AAGAAAGATT TGGACCGCAT GCAGCCTCCT 1519
*
445
TCTAGCAGCT GTTAGGTTGT ATTGTTATTT ATGGATGAGT TTGTAGAGCG GTGGGGTTAA 1579
CTTTAACAGC AAGGAAGCTC TGGTGACCAG GCTGATTGGC TTAGAAGTTA TGGGAACCCC 1639

TTGAAAGGGT CAGGGTAAA TATATTCAG TTGTTTATT AGTGATTATC TTGTGGTAA 1699
CTTATACGAA TGCAAATCAT TCTATGCAGT TTTTCTTCGT CCCACTTGT TTGGCTTCTC 1759
TATTGCTAGT GTACATATCT CTTCAAACAT GTACTAAATA ATGCGTGTG CTTCAAAGAA 1819
GTAACCTTTA TTAAAAAAA AAAAAAAAC TCGAG 1854

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 445 amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: Protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Ala Arg Ile Ser Cys Asp Leu Arg Phe Leu Leu Ile Pro Ala Ala
1 5 10 15
Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln Thr Gln Ser Gln
20 25 30
Tyr Ala Asp Arg Leu Ser Ser Ala Ile Glu Ser Glu Asn His Cys Thr
35 40 45
Ser Gln Met Arg Gly Leu Ile Asp Glu Val Ser Ile Lys Gln Ser Arg
50 55 60
Ile Val Ala Leu Glu Asp Met Lys Asn Arg Gln Asp Glu Glu Leu Val
65 70 75 80
Gln Leu Lys Asp Leu Ile Gln Thr Phe Glu Lys Lys Gly Ile Ala Lys
85 90 95
Leu Thr Gln Gly Gly Gln Met Pro Val Ala Ala Val Val Met Ala
100 105 110
Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys Ser Val Leu Thr
115 120 125
Tyr Gln Thr Pro Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln Asp
130 135 140
Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser Tyr Asn Gln Leu
145 150 155 160
Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val Val Thr Glu Arg Pro
165 170 175
Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp Ala
180 185 190
Leu Asp Gln Leu Phe Tyr Lys His Lys Phe Ser Arg Val Ile Ile Leu
195 200 205
Glu Asp Asp Met Gln Ile Ala Pro Asp Phe Phe Asp Tyr Phe Glu Ala
210 215 220
Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met Ala Ala Ser Ser

240

235

230

225 Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp Pro Tyr Ala Leu
245 250 255

Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Lys Arg Ser
260 265 270

Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp Asp
275 280 285

Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln Phe Ile Ala Pro
290 295 300

Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu Gly
305 310 315 320

Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val Thr
325 330 335

Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr Glu Gly Asn Tyr
340 345 350

Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg Pro Ile Gln Gly
355 360 365

Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp Asp Arg Ile
370 375 380

Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala Gly Glu Phe Gly
385 390 395 400

Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr Ala Tyr Lys Gly
405 410 415

Val Val Val Phe Arg Ile Gln Thr Thr Arg Val Phe Leu Val Gly
420 425 430

Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser *
435 440 445